# The User-Subjective Approach to Personal Information Management Systems Design: Evidence and Implementations

**Ofer Bergman**
*Tel Aviv University, Sharet Building, Ramat Aviv, Tel Aviv 69978, Israel. E-mail: oferb@post.tau.ac.il*

**Ruth Beyth-Marom**
*Open University of Israel, 108 Ravutski St. POB 808, Raanana 43107, Israel, E-mail: ruthbm@openu.ac.il*

**Rafi Nachmias**
*Tel Aviv University, Sharet Building, Ramat Aviv, Tel Aviv 69978, Israel, E-mail: nachmias@post.tau.ac.il*

**Personal Information Management (PIM) is an activity in which an individual stores personal information items to retrieve them later. In a former article, we suggested the user-subjective approach, a theoretical approach proposing design principles with which PIM systems can systematically use subjective attributes of information items. In this consecutive article, we report on a study that tested the approach by exploring the use of subjective attributes (i.e., project, importance, and context) in current PIM systems, and its dependence on design characteristics. Participants were 84 personal computer users. Tools included a questionnaire (N = 84), a semistructured interview that was transcribed and analyzed (n = 20), and screen captures taken from this subsample. Results indicate that participants tended to use subjective attributes when the design encouraged them to; however, when the design discouraged such use, they either found their own alternative ways to use them or refrained from using them altogether. This constitutes evidence in support of the user-subjective approach as it implies that current PIM systems do not allow for sufficient use of subjective attributes. The article also introduces seven novel system design schemes, suggested by the authors, which demonstrate how the user-subjective principles can be implemented.**

Personal Information Management (PIM) is an activity in which an individual stores personal information items to retrieve them later. PIM is performed in a physical environment (e.g., an office), with mobile devices (e.g., mobile phones and PDAs), and by using personal computers. The information items used in personal computers include documents, e-mail, Web Favorites, tasks, and contacts. Following Soergel's (1985) model, Barreau (1995) listed five PIM activities: acquisition, organization and storage, maintenance, retrieval, and output. In the personal computer environment, these activities are supported by PIM systems such as the operating system, mailbox, and browser.

Despite the fact that PIM is a fundamental aspect of computer-based activity and millions of computer users manage their personal information on a daily basis, there is surprisingly little research on the subject (Whittaker, Terveen, & Nardi, 2000); however, in recent years, the topic has begun to attract increasing scientific attention (Bergman, Boardman, Gwizdka, & Jones, 2004; Jones & Bruce, 2005; Teevan, Jones, & Bederson, 2006). Many PIM-related studies report on users' problems with the classification and retrieval of their personal information as well as their dissatisfaction with these processes. Problems are not restricted to a specific PIM system and are reported for digital files (Barreau, 1995; Barreau & Nardi, 1995), e-mail (Whittaker & Sidner, 1996), and Web Favorites (Abrams, Baecker, & Chignell, 1998; Jones, Bruce, & Dumais, 2001). Boardman, Spence, and Sasse (2003), who studied PIM within and across these systems, made the following observation:

> We were often surprised at the vehemence expressed regarding PIM-related problems, and have coined the term bugbear for recurring problems that frequently or seriously affect users. Since PIM is an ongoing and often repetitive everyday activity, we found that even relatively minor bugbears can build up and have a negative impact on productivity and/or user experience. (p. 618)

Received July 23, 2006; revised March 5, 2007, July 18, 2007; accepted July 18, 2007

 ● Published online 26 October 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20738

It seems that many users blame themselves for being disorganized (Bellotti & Smith, 2000), though part of the problem can be ascribed to faults in PIM design.

Kwasnik's (1991) research shed some light on this design problem. She analyzed the descriptions of eight faculty members who were asked to describe how they organized physical documents in their offices. She found that a minority (30%) of the attributes described were document-related (e.g., author, form, topic, title) while the majority (70%) were attributes related to the interaction between the user and the information (e.g., situation attributes, disposition, time, cognitive state). These results suggest that users base their natural data organization more on subjective attributes than on general "objective" ones.

In a former article published in this journal, titled "The User-Subjective Approach to Personal Information Management Systems" (Bergman, Beyth-Marom, & Nachmias, 2003), we suggested a theoretical design approach aimed to improve PIM systems design. The user-subjective approach recommended design principles with which PIM systems can systematically make use of subjective attributes. The purpose of the present, consecutive article is twofold: (a) to provide empirical evidence that might shed light on the questions regarding the use of subjective attributes in current PIM systems, and (b) to present seven novel system design schemes, suggested by the authors, to demonstrate how the user-subjective approach can be implemented.

## The User-Subjective Approach

PIM system designers may be partly to blame for users' dissatisfaction and complaints about their information management, as they fail to recognize the unique requirements of such systems. In contrast to other information management systems, which serve many different users, each PIM system serves only one person. In other systems, information is stored and organized by information professionals (e.g., librarians, Web site developers) for users who retrieve information according to their needs. To cater to the needs of different users and facilitate information retrieval, information professionals use general and objective attributes of the information for its organization. PIM systems are unique in that the same person who stores the information and decides on its organization is the one who later retrieves it. The user-subjective approach takes advantage of this unique feature and suggests that PIM systems should make use of subjective, user-dependent attributes. The system should capture these subjective attributes when the user interacts with the information item (either automatically or by using direct manipulation design) to help the user retrieve that item later.

### *Subjective Attributes*

Attributes of an information item are variables which describe it, and in so doing, add value to it (Taylor, 1986). We differentiate between objective and subjective attributes. Objective attributes are attributes that are user-independent, in the sense that an external observer can infer these attributes directly from the information item, without observing the user's actions. Such objective attributes are the item's format, size, and date. In contrast, subjective attributes are user-dependent and cannot be inferred directly from the information item. Instead, they often can be derived from the user-information interaction. The user-subjective approach identifies three subjective attributes: the project to which the item belongs, its importance to the user, and the context in which the item is used.

*Project attribute.* Information items in PIM often are classified under projects with which the users are involved (Jones, Phuwanartnurak, Gill, & Bruce, 2005). "Project" is a subjective attribute of an information item chosen by the PIM user; thus, the same information item can be classified under different projects for different users. For example, a person attending the ASIST 2006 Conference can classify the URL of the conference hotel in a folder designated "ASIST 2006 Conference" within his or her Web Favorites; however, the same URL will be placed in a "Honeymoon" folder for another person planning to visit the hotel for that purpose.[1]

*Importance attribute.* Importance is user-dependent and thus subjective: An information item can be very important to one person and completely unimportant to another. It also can be important (i.e., relevant) to a user today, but unimportant (i.e., irrelevant) to that same user next month. One measure of importance can be derived from the user-information interaction: Recently used information items are probably judged as more relevant to the user than are information items that have not been used for a long time. The former are thus more likely to be retrieved again in the near future.

*Context attribute.* Context is a subjective attribute because different people interact with the same information items in different contexts. The term *context* has been used extensively in the literature and may have different meanings (Chalmers, 2004; Dey, 2001; Dourish, 2004). The user-subjective approach refers to four context attributes of an information item: external, internal, temporal, and social. The *external context* of an information item refers to the other items that the user dealt with while interacting with a specific information item; *internal context* relates to the user's thoughts while interacting with the information item; *temporal context* pertains to the state in which the user left the information item when she or he last interacted with it, and to his or her working plans regarding that information; and *social context* refers to other persons relating to the information item, such as other people who collaborate with the user regarding that information item.[2]

[1]The term 'project' replaces the term 'subjective topic' used in our previous paper for this attribute, as it is commonly used in the literature.

[2]Project and importance are also context-related, and were presented separately for the sake of convenience.


DOI: 10.1002/asi

The three attributes (i.e., project, importance, and context) are user-dependent, and one cannot infer them just by looking at the information item; however, they can be deduced from the user-information interaction.

*User-Subjective Principles*

The user-subjective approach suggests three design principles, each of which takes into account one of the subjective attributes mentioned earlier. These principles are only occasionally implemented in current PIM systems design. We will describe these principles and their sporadic implementation.

*Subjective project classification principle.* The subjective project classification principle suggests that design should allow all information items related to the same project to be classified under the same category regardless of their technological format. Although project-based classification has been encouraged in experimental systems (Bellotti & Smith, 2000; Dourish, Edwards, LaMarca, & Salisbury, 1999; Freeman & Gelernter, 1996; Jones, Munat, & Bruce, 2005; Kaptelinin, 2003; Karger & Quan, 2004), current PIM system design discourages such classification. At present, it encourages users to classify their information items within their format-related hierarchy: Documents are stored in one folder hierarchy (e.g., in My Documents), e-mail in a separate mailbox hierarchy, and favorite Web sites in yet another browser-related hierarchy (These will be referred to as the three hierarchies.) Thus, information items related to the same project, but based on different formats (e.g., documents, e-mail, or Web Favorites), are retrieved from different collections stored in separate locations (i.e., hierarchies), causing "the project fragmentation problem" (Bergman, Beyth-Marom, & Nachmias, 2006). The only exception to this fragmentation problem in current PIM systems design is in the case of documents: Different kinds of documents (e.g., Word, Excel, and PowerPoint documents) are all classified under one hierarchy.

*Subjective importance principle.* The subjective importance principle suggests that the importance of information should determine its degree of visual salience and accessibility: Highly important information items should be highly visible and accessible as they are likely to be retrieved; information items of lower importance should be less visible so as not to distract the user. In current PIM systems, there are several ways to make information items of high importance visible and accessible, such as saving them (or creating shortcuts to them) on the desktop, using the recently used items, and the history options. However, current systems offer no way of making information items of low subjective importance less visible.[3] As a result, folders are often cluttered with information items of low importance which compete for the user's attention (Jones, 2004). Moreover, users have difficulty "seeing the forest for the trees" when, for example, they have to identify a new version of a document in a list of many old and new documents.

*Subjective context principle.* The subjective context principle suggests that information should be retrieved and viewed by the user in the same context in which it was previously used, to bridge the time gap between these two events. In recent years, there has been a strong tendency to develop context-aware systems (Chalmers, 2004; Dey, 2001; Dourish, 2004). Recent PIM systems allow the user to mark the state of the information item in the process of the user's work with it (i.e., the temporal context) by using flags (in MS Office applications) and tags (Cutrell, Robbins, Dumais, & Sarin, 2006), and some PIM systems (e.g., the Macintosh operating system) allow the user to annotate an entire information item (i.e., the internal context). However, current information systems do not allow the user to restore the environment in which the user had previously worked with the information item (i.e., the external context). This is true for information items that obviously relate to each other, such as an e-mail and its attachment, and also for information items that might relate to each other as they were open at the same time. In addition, current systems do not offer a way to maintain and use the social context of information items—the list of people involved with the same information item.

The present research tested the user-subjective approach and its design principles.

## Research Questions

The purpose of the present research was to explore the use of subjective attributes in current PIM systems, and its dependence on design characteristics. This was done regarding each of the three subjective attributes and principles. Following are the general questions and the different empirical measures we chose for the exploration of each question.

1. Questions regarding the subjective project classification principle: Do personal computer users tend to work with their information items according to the items' formats (as suggested by current PIM systems design) or according to the projects with which the items are associated (as suggested by the user-subjective approach)?
   - 1.1. How do users tend to talk about their information organization—in terms of technological formats, or in terms of projects?
   - 1.2. To what extent do recently used files, e-mail, and Web pages relate to the same projects?
   - 1.3. How much overlap is there in the folder names between the three folder hierarchies (i.e., documents, e-mail, and Web Favorites)?
   - 1.4. Do users tend to classify their information according to format or project?
   - 1.5. To what extent does interface design affect the format heterogeneity of information items classified under the same folder?

[3]Two exceptions to this are the ability MS Windows gives the user to set aside unused desktop items, and the ability MS Outlook gives the user to indicate that a task is of low priority.


DOI: 10.1002/asi

2. Questions regarding the subjective importance principle: Do personal computer users relate to the importance of information items while storing and/or retrieving them?
   2.1. Do users tend to talk about their information organization in terms of its subjective importance?
   2.2. To what extent do users report using importance attributes while working with their PIM systems?
3. Questions regarding the subjective context principle: Do personal computer users tend to relate to the context of information items?
   3.1. Do users tend to talk about their information organization in terms of its context?

As current PIM systems lack context-related features, we could not explore to what extent users use these features when working with their PIM system.

## Method

*Participants*

Participants were 84 computer users (23 men, 61 women) from Israel, varying in age (21–57 years, $M = 30.44$, $SD = 7.86$), occupation, and degree of computer literacy. Their experience with personal computers ranged from 0.5 to 22 years ($M = 8.83$, $SD = 4.99$). Sixty-nine participants used a personal computer with an MS Windows operating system, 3 used a Macintosh, and 4 used both platforms (Eight participants did not specify their platform.)

*Tools*

*Questionnaire.* All 84 participants completed a questionnaire that included items on how they managed their documents, e-mail, and Web Favorites. They were specifically asked questions regarding the three subjective attributes to explore the aforementioned research questions. Most questions in the questionnaire were measured on a Likert-type scale of 1 (not at all) to 5 (to a very large extent). Some background information also was collected.[4]

*Interview.* Twenty participants[5] were interviewed in a semi-structured technique known as the *guided tour*: The participant guides the interviewer through her or his computer, focusing on the way she or he organizes information and retrieves it (for previous research that used this technique, see Barreau, 1995; Boardman & Sasse, 2004; Kwasnik, 1991; Malone, 1983). We gave the participants the following instructions: "Please show me how you manage your personal information. That is, how you organize your files, emails and Web favorites and how you retrieve an information item when needed." Each interview was recorded, transcribed, divided into paragraphs, and analyzed in the following way: We defined six dichotomous (yes/no) variables, and two independent judges graded each paragraph on each variable. Three of the variables concerned subjective attributes: Was there a reference to project (i.e., topic or content that the user is involved with), to importance (both high and low), and to context (i.e., external, internal, temporal, and social)? The other three variables concerned objective attributes: Was there a reference to format (including implicit reference such as "pictures" and "presentations"), to size (e.g., "I try to get rid of heavy files"), and to date (e.g., "This is from 2004")? The judges graded only spontaneous paragraphs, excluding those in which the participant answered an interviewer's question. We only included the 566 judgments on which both judges agreed (88% of all judgments) in the content analysis.

*Screen captures.* Two types of screen captures were taken and printed out: information items and folders. Information item screen captures are recently used information items of the three hierarchies: documents, e-mail, and Web pages used the day before the interview (using Recent Documents, and Web History functions). Folder screen captures are screen captures of the three folder hierarchies at root level.

To increase convergence validity, we employed a multi-operational method: Whenever possible, questionnaires, interviews, and screen captures of participants' computers were used to investigate the same principle (1, 2, or 3). Table 1 presents the research questions and the tools chosen to explore them.

## Results

Next, we present results according to the research questions that tested the subjective project-classification, importance, and context principles.

*Testing the Subjective Project Classification Principle*

The general research question regarding the subjective project-classification principle was: Do personal computer users tend to work with their information items according to the items' formats (as suggested by current PIM systems design) or according to the projects with which the items are associated (as suggested by the user-subjective approach)? We tested this question using five measures.

TABLE 1. Research questions and tools.

| Research question | Research tool |
|---|---|
| 1.1 | Interview |
| 1.2 | Screen captures–items |
| 1.3 | Screen captures–folders |
| 1.4 | Screen captures–folders |
| 1.5 | Questionnaire |
| 2.1 | Interview |
| 2.2 | Questionnaire |
| 3.1 | Interview |

[4]To receive the full questionnaire, please contact the first author.

[5]After interviewing 20 participants, we felt that the research had reached saturation.


DOI: 10.1002/asi

1.1 How do personal computer users tend to talk about their information organization—in terms of technological format or in terms of projects?

In the interviews, the participants referred to item format only occasionally. They mostly referred to projects when asked about the way they organize their information: Researchers talked about their research projects, a journalist about articles he wrote, a musician about CDs he composed, illustrators about books they illustrated, a multimedia designer about titles she designed, a movie producer about videos he produced, and so forth. The tendency to refer to projects more than to format was evident in the content analysis; for each interview, two measures were calculated: the percentage of paragraphs in which at least one project was mentioned and the percentage of paragraphs in which a technological format was referred to at least once. An average of 71% of the paragraphs contained references to projects while only 28% of them referred to technological formats (see Table 2, Lines 1 and 2). This huge difference was significant, $t\,(19) = 8.88$, $p < .01$. The following is a typical example of a paragraph containing several references to projects taken from an interview with an artist:

> Here I have an archive of things that I downloaded from the Internet that I keep for [visual] references. It keeps on getting bigger and bigger. This is a business card I started making for myself. 'Comme il faut' is for a new [art] exhibition I'm making in a café by that name . . . ."

1.2 To what extent do recently used files, e-mail, and Web pages relate to the same projects?

To answer this question, we used screen captures of participants' documents, e-mail, and Web pages that were used the day before the interviews took place (using Recent Documents, and Web History functions). These were printed and analyzed: Next to each information item (regarding documents, e-mail, and Web pages), participants wrote the project to which it related. We defined an overlapping information item as one that has another information item relating to the same project located in at least one different format collection (e.g., an e-mail and a document that referred to the same project). Overlap was measured by the percentage of overlapping items among all previous-day items for each participant. As can be seen in Table 2, Line 3, there is an average overlap of 56% of the information items. Thus, when working on a project, about half of the time, participants retrieve items from different format hierarchies.

1.3 How much overlap is there in folder names between the three folder hierarchies?

To answer this question, we used folder screen captures of the three folder hierarchies at root level. The printouts showed 968 folders: 544 document folders, 261 e-mail folders, and 163 Web Favorites folders. Overlapping folders were defined as folders of different hierarchies with the same name, or those the user explicitly mentioned as relating to the same project. For example, a Comics artist had a folder with the name of a book (on which he was working at the time), both in his files and in his mailbox directories. The results show that on average, 20% of the folders overlapped (see Table 2, Line 4). In other words, nearly one fifth of the folders had another folder relating to the same project in a different hierarchy.

TABLE 2. Testing the subjective project classification principle.

| | *M* (*SD*) |
|---|---|
| Paragraphs that mentioned: | |
| 1. projects | 71 (16) |
| 2. format | 28 (15) |
| Project overlap of: | |
| 3. information items in different formats | 56 (33) |
| 4. folder in different hierarchy | 20 (19) |
| Classification according to: | |
| 5. project name | 80 (12) |
| 6. format name | 6 (7) |

1.4 Do users tend to classify their information according to format or to projects?

The 968 folders in the three different hierarchies mentioned earlier were classified according to their names (by project, by format, by name of person, and by other names). For each participant, the percentage of the four types of names was calculated in each of the three hierarchies. The average proportion of project folder names was 80%[6] while the average proportion of format folder names was only 6% (see Table 2, Lines 5 and 6). This enormous difference was significant, $t\,(19) = 19.12$, $p < .01$.[7] Other categories of folder names were names of people ($M = 7\%$, $SD = 6\%$), and other names ($M = 7\%$, $SD = 7\%$). These results indicate that users tend to classify their information according to projects more than to formats.

1.5 To what extent does interface design affect the format heterogeneity of information items classified under the same folder?

My Documents is the default storage location for most documents regardless of their file format (e.g., Word, Excel). This interface encourages the user to classify documents of heterogeneous formats under the same project. In contrast, e-mail and Web Favorites are each stored, by default, in separate designated locations. While the system enables storing e-mail and Web Favorites together with documents,

[6]Results were consistent across the three hierarchies: the average proportion of project names was over 70% for each.

[7]One possible explanation for this is that users don't need to organize their files according to format because they are able to search according to their format by specifying the file extensions.


DOI: 10.1002/asi

the system interface discourages such activity (E-mail stored as documents cannot be replied to, and most participants did not know that Web Favorites could be stored with files.)

To what extent does this difference in interface design affect users' behavior? In the questionnaire, participants were asked about their storage habits. Specifically, they were asked about the extent to which they saved documents of different formats but relating to the same project in one project folder (an activity which *is* encouraged by the system design), and about the extent to which they saved e-mail, Web Favorites, and Web pages with documents that related to the same project (an activity which *is not* encouraged by the system design). Participants indicated, on a Likert-type response scale of 1 (not at all) to 5 (to a very large extent), that they often mix documents of different formats in the same folders ($M = 3.35$, $SD = 2.32$), but they seldom store e-mail, Web Favorites, and Web pages outside their default locations (e.g., storing them in file folders) ($M = 1.67$, $SD = 1.63$). A paired $t$ test showed a significant difference between results, $t(83) = 11.48$, $p < .01$, indicating that users' tendency to store information items of different formats in the same folders may be determined by interface design.

### *Testing the Subjective Importance Principle*

The general research question testing the subjective importance principle was: Do personal computer users relate to the importance of information items while storing and/or retrieving them? We explored this question using two measures.

2.1 Do users tend to talk about their information organization in terms of its subjective importance?

In the interviews, participants tended to talk about their personal information organization both in terms of high importance (e.g., "Come to think of it, I made shortcuts [pointing at his desktop shortcuts] to most of the folders that are important for me") and in terms of low subjective importance (e.g., "These are sketches, they are not really important, just drafts" and "Once every 2–3 days I go through [unimportant] email to send it to the trash bin"). Occasionally, during the interviews, participants also deleted information items saying that they "no longer needed them." The percentages of paragraphs containing spontaneous references to high subjective importance as well as to low subjective importance were calculated for each participant, as part of the content analysis. On average, 22% of the paragraphs contained high-importance references ($SD = 17\%$), and 29% contained low-importance references ($SD = 14\%$).

2.2 To what extent do users report using importance attributes while working with their PIM systems?

*High importance.* To explore whether participants make highly important items more visible, the questionnaire asked about their retrieval habits. Concerning files, they were asked to divide 100% of their document retrieval activity between five options: direct navigation to the relevant folder and file, search, desktop shortcuts, recently used documents lists, and "others." On average, participants answered that they retrieved 18% of their files by using desktop shortcuts and 12% by using recently used documents lists (see Table 3, Lines 1 and 2). In addition, participants were asked: "When you are looking for a file inside a folder, you can sort the files in chronological order so that the files you used recently would appear at the top of the list. To what extent do you use this option?" The participants' average response, on a scale of 1 (not at all) to 5 (to a very large extent), was 1.81 ($SD = 2.3$). When asked about refinding Web pages that they had seen before, participants estimated that they used the Web Favorites option for 32% of the retrievals, on average, and the History option for 13% of their retrievals (see Table 3, Lines 3 and 4). The remaining 55% included writing the address, using the search engine, or links from other Web sites. Thus, results indicate that participants used the tools that the current design offers them to make highly important items more visible and accessible.

*Low importance.* As current design does not offer a designated option to make files of low subjective importance less visible, the questionnaire asked participants about various alternative ways of doing this. The results show that 40% of the participants indicated that they transferred at least some of such items to a general archive folder, 61% moved them to an external memory such as a CD, 32% created a new folder for the same purpose and used the old one as an archive, and 24% of the participants indicated that they created an archive folder within the original folder (see Table 3, Lines 5–8). Altogether, 79% of the participants used one or more of these alternative ways to make at least some of their low subjective importance files less visible.

### *Testing the Subjective Context Principle*

The general research question regarding the subjective context principle was: Do personal computer users tend to relate to the context of information items? Investigation of this question was based on content analysis of the interviews. Each paragraph was coded as either having or not having a

TABLE 3. Options chosen by users to increase or diminish item visibility.

| Increase item visibility by using: | *M* (*SD*) |
|---|---|
| 1. Desktop shortcuts | 18 (24) |
| 2. Recently used documents lists | 12 (19) |
| 3. Web favorites option | 32 (29) |
| 4. Web history option | 13 (18) |
| **Diminish item visibility by using:** | **% who chose each option** |
| 5. A general archive folder | 40 |
| 6. An external memory | 61 |
| 7. An old folder as an archive | 32 |
| 8. Archive folder within the original folder | 24 |


DOI: 10.1002/asi

TABLE 4. References to context attributes.

| Contextual attributes | *M* (*SD*) |
|---|---|
| External context | 7 (10) |
| Internal context | 1 (4) |
| Temporal context | 29 (21) |
| Social context | 25 (16) |
| At least one attribute | 50 (23) |

reference to external, internal, temporal, and social context. Table 4 presents the percentages of paragraphs containing at least one reference to each of the context attributes.

*External context.* When during the interview participants mentioned storing together two or more information items that are used simultaneously so that they could easily be retrieved together, that paragraph was coded as referring to external context. For example, YP, who is a Comics artist, has two kinds of folders: text folders and image folders. When creating a comics story, he reported that he copied the relevant text folder to the relevant image folder because "I need it in order to work there." Results of the content analysis show that an average of 7% of the paragraphs referred at least once to external context.

*Internal context.* The interviewees rarely mentioned writing down their thoughts about their information items (perhaps because most of their system designs do not allow for such action). An example of a reference to an internal context can be seen in OD's (a research student) interview. She reported saving her thoughts about her results in the same folder in which she kept her results "so that later when I put it all together I'll know on which results I based each of my conclusions." The content analysis shows that only 1% of the paragraphs on average referred to the internal context.

*Temporal context.* While describing how they organized their personal information, participants often talked about their plans regarding their information items; they described what they had done with the information item in the past, the current state of the information item, and what they intended to do with it in the future. The content analysis reveals that on average, 29% of the paragraphs in the interviews referred to temporal context. For example, in an interview with the editor of a musical journal who works on a Macintosh, he said: "Each time I edit a new journal edition, I give its folder this icon [showing the icon]. Then, when I'm done working on it, I return to the standard folder icon." In this way, he can identify at a glance the folder he is currently working on.

*Social context.* When information items were sent to interviewees from other people, or when an information item was produced collaboratively with others, participants often mentioned these social contacts when referring to the relevant information item. The content analysis reveals that on average, 25% of the paragraphs referred to social context. This reference was prominent mainly when participants described their e-mail organization, but also when they guided the interviewer through other collections. An example of this is a medical doctor who said, "In here I have a folder for [the files of] each of my patients."

Overall, the content analysis shows that on average, 50% of the paragraphs referred to at least one contextual attribute.

## Discussion

Overall, the results indicate that participants tend to talk about their personal information management in terms of subjective attributes, and use these attributes when the design allows them to; however, when the PIM system design does not allow them to use these subjective attributes, they either refrain from using them (e.g., participants rarely store e-mail and Web Favorites together with documents relating to the same project) or find their own alternative way to use them (e.g., by using various ways of archiving documents of low subjective importance). Thus, though needed, current PIM systems do not allow for sufficient use of subjective attributes, and it is recommended that they should be designed to allow for broader and more systematic use of subjective attributes.[8]

Evidence regarding each of the user-subjective principles will be discussed next, followed by examples of their implementation. These implementations are design schemes, suggested by the authors, aimed to improve existing PIM systems. Each will be developed and tested in future research.

### *Subjective Project Classification Principle: Evidence and Implementations*

Regarding the subjective project classification principle, results indicate that personal computer users tend to refer to their information items according to the project with which the items are associated, more than in terms of their formats. Though they use different information-item formats (documents, e-mail, and Web Favorites) while working on the same project, they usually save those in the three format-related hierarchies. Occasionally, the folders in which those items are stored (in the different hierarchies) are named similarly according to the relevant project.

Results also indicate that personal computer users tend to store project-related information items of different formats in one project folder and retrieve them from there, when the interface design encourages it (i.e., with different document formats);[9] however, they store project-related information items in different folder hierarchies (documents,

[8]A possible limitation of these results is that most of them were based on participants' reports and only a minority on users' actual behavior.

[9]See "To what extent does interface design affect the format heterogeneity of information items classified under the same folder?" in the results section.


DOI: 10.1002/asi

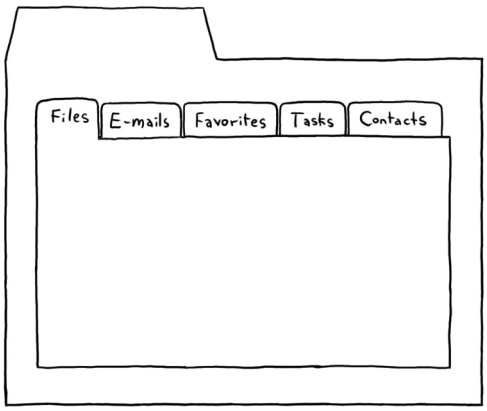


FIG. 1. *ProjectFolders* scheme.

e-mail, and Web Favorites) when the design encourages such fragmentation.

Project fragmentation occurs when a user who is working on a single project stores information items relating to that project in separate format-related collections from which he or she also retrieves them. The results reveal the problem: Users tend to relate to their information items in a certain way (i.e., according to projects); however, current design discourages them from storing and retrieving them in the same way (by suggesting format-related storage).

The subjective project classification principle (which suggests that design should allow all project-related information items to be classified together regardless of their technological format) can be implemented in various ways to address the project fragmentation problem. In this article, we suggest two such design schemes: *ProjectFolders*, a more radical storage solution, and *MyProjects*, a more lightweight retrieval solution. In addition, we suggest *ProjectField*, a feature designed to enable application of the subjective project attribute in e-mail.

*ProjectFolders.* The most straightforward way to address the project fragmentation problem by using the subjective project classification principle is to store all project-related items regardless of their format in a single folder hierarchy. In *ProjectFolders* system design, the user will be able to store all project-related documents, e-mail, and Web Favorites tasks and contacts in a single folder, with tabs separating them (Figure 1) (Bergman et al., 2006).[10] This will allow users to work in the context of their projects and retrieve all their project-related items from a single location. Earlier attempts to address the project fragmentation problem assumed it was inevitable and suggested solving it by using search (Dourish et al., 1999; Dumais et al., 2003; Freeman & Gelernter, 1996) or by using an additional structure (Bellotti & Smith, 2000; Kaptelinin, 2003; Karger & Quan, 2004). ProjectFolders attacks the problem itself by avoiding the creation of fragmentation in the first place. It is a radical solution to the project fragmentation problem in the sense that it requires changes in users' storage habits. Moreover, the present study found a folder-name overlap of only 20% between the three hierarchies, thus supporting the claim that using a single folder hierarchy may be questionable, as users tend to use different strategies to manage the three hierarchies (Boardman & Sasse, 2004). However, interface design often determines users' preferences and strategies; changing the interface may change user behavior and improve usability. In the early 1990s, when each document application suggested a separate storage location, users may have been shown to employ different storage strategies for the different document formats; however, now that current systems offer a single storage location for all documents (e.g., My Documents), users tend to store project-related documents of different formats in the same folders, as indicated by our data. Thus, ProjectFolders design may lead to a similar change for all project-related information items.

*MyProjects. MyProjects* is a second design scheme aimed to address the project fragmentation problem by allowing the user to retrieve all the information items (i.e., documents, e-mail, and Web Favorites) related to the same project in a single menu (Figure 2). It does so by creating a common index for folders with the same names in the three folder hierarchies. The tool automatically creates an index which points at folders in the three indexes. When two or more of the folders in the three hierarchies have the same name (e.g., a folder named "Chemistry" both in My Documents and in the mailbox), the tool creates a common index for these folders. This common index allows the tool to present recently used information items contained in these folders when this project is selected. *MyProjects* is minimal both in what it requires from users and in its ambition to solve the project fragmentation problem: The users do not need to make any changes in their storage habits. However, the tool is effective only for the minority of cases when there is a name overlap between folder hierarchies.

*ProjectField.* When retrieving an e-mail message, it is often difficult to determine the project to which the message belongs only by observing its subject. This problem is often mentioned by groups of workers who correspond with one

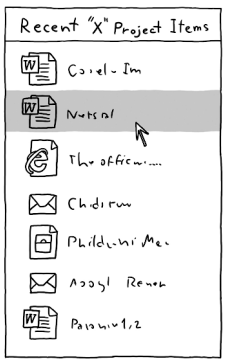


FIG. 2. *MyProjects* scheme.

[10]Technically, email messages will not be stored in these folders, only pointers to the emails in the mailbox database; however, this will be transparent to the user.


DOI: 10.1002/asi

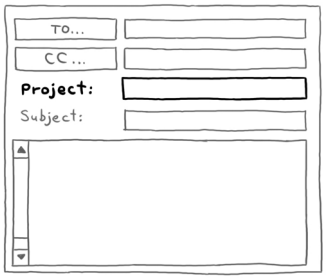


FIG. 3. *ProjectField* scheme.

another regarding several projects at once. *ProjectField* is an additional field in the e-mail message that indicates the project to which it belongs (Figure 3). This also allows the user to sort messages by projects. When a document is sent by e-mail, *ProjectField* will automatically take on the name of the document folder, similarly to the way the e-mail subject takes on the name of the document. A possible reason for complications with this feature is that for it to work properly, all correspondents need to use the same project names.

### *Subjective Importance Principle: Evidence and Implementations*

When testing the subjective importance principle, results show that participants tend to talk about their information items in terms of both high importance and low importance. Participants also tend to use system design features that support easy retrieval of highly important information items (e.g., Desktop Shortcuts). However, as the system does not offer a way of making items of low importance less visible, most users find their own alternative ways of doing so. This may suggest that system design also should support features for making information items of low importance less visible.

Why is it important to deal with information items of low importance? Jones (2004) claimed that the decision whether "to keep or not to keep" information for future usage is prone to two types of costly mistakes: Information not kept or deleted may be unavailable later when it is needed; however, if unimportant information items are kept, they compete for the user's attention and may obscure information important to the current task. Furthermore, the decision whether to keep or delete information takes the user's time and cognitive resources (Bergman et al., 2003). To avoid these problems, the subjective importance principle offers an intermediate option between "keep" and "delete" ones. It suggests that information items of low importance will be less visible without denying the user future access to that information item. This option has the advantage of the "delete" option (less competition for attention) as well as the advantage of the "keep" option (access to information item in case it is unexpectedly needed in the future), and saves the time and energy it takes to decide what to do with the item in question.

When designing features for making information items of low importance less visible, it is important that they should not be removed from the folders in which they were originally stored. In this way, the items can be retrieved in their original context, and the user will not have to remember another storage place associated with them. One way of doing this is by changing the default sort of the information items from alphabetical order to date order, so that documents that have not been retrieved for a long time would appear at the bottom of the list. Two alternative design schemes that implement the importance principle are *Gray Area* and *Old'nGray*.

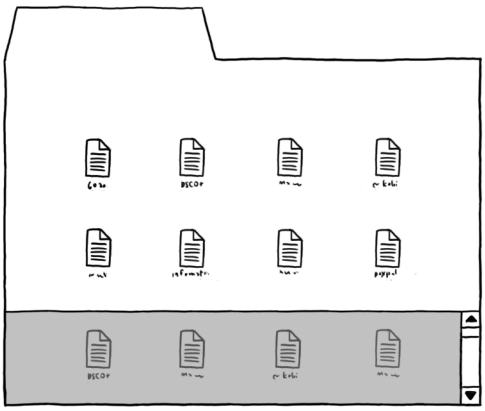

FIG. 4. *GrayArea* scheme.

*GrayArea. GrayArea* is an additional folder feature that allows the users to drag information items of low importance to a designated location at the bottom of a folder (Figure 4). *GrayArea* uses the term "gray area" because it is a storage place for information items that users are in doubt as to whether to delete. Information items in that area are presented in a smaller font, are concealed by the area's gray background, and are presented in a small space (If information items exceed the gray area, a scroll bar is used.) Items within the *GrayArea* also may be compressed to free disc space.

*Old'nGray.* When retrieving a document, users face a constant dilemma: Which version is the latest one? To answer that question, users usually go through a time-consuming process of changing the folder's file listing from alphabetical to chronological order, and compare dates of versions. The situation is complicated further as different versions of the document may have different names and be stored in different folders. *Old'nGray* is an operating system level feature that solves this problem by automatically fading to gray old versions of documents when a new version is stored on the computer. In this way, the user can tell at a glance the latest version from the older ones (Figure 5). *Old'nGray* also will allow the user to go from the older version to the latest one, even if it is stored under another name or in another folder on that computer.

The use of *GrayArea* and *Old'nGray* is not limited to retrieval from folders; it also can be applied to retrieval through search. Currently, search results often include information items of low subjective importance. For example, considerable time may be spent on distinguishing the latest version of a


DOI: 10.1002/asi

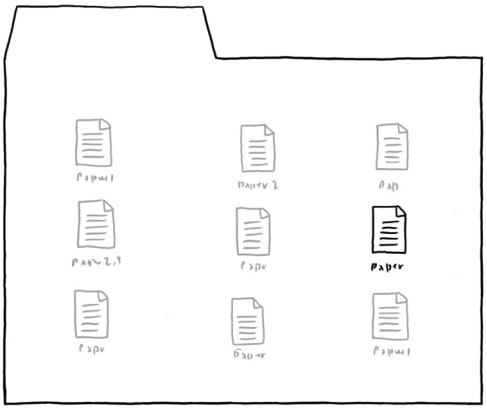


FIG. 5. *Old'nGray* scheme

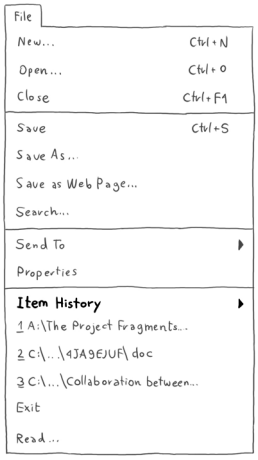


FIG. 6. *ItemHistory* scheme.

document from the older ones stored in other folders (e.g., an archive folder), as queries typically search across folders (Jones, Karger, et al., 2005). When the user uses search using these applications, the information items marked by *GrayArea* and *Old'nGray* will be distinguished from the rest of the results in the list by their color. Moreover, the search engines also could use this low subjective attribute and place these information items at the bottom of the search results list.

### *Subjective Context Principle: Evidence and Implementations*

When testing the subjective context principle, results show that participants tended to talk about their personal information organization in terms of external, temporal, and social context, but not in terms of internal context. The notion that information should be understood in its context is widely accepted in the field of information science (Medlin & Schaffer, 1978; Saracevic, 1999; Schamber, 1994) and in the specific field of PIM research (Barreau, 1995; Kwasnik, 1991; Malone, 1983); however, current PIM systems make only limited use of context-related features in their design. Two examples of context-dependent features that can be added are shown next.

*ItemHistory (external context).* In his 1945 article, "As We Might Think," Vannevar Bush envisioned a machine he termed "Memex," which automatically indexed information items sequentially viewed by the user in associative trails to allow her or him to return to these information items in the future (Bush, 1945). More than 60 years after the article was written, current PIM systems still do not allow the user to follow a trail from one information item to other information items that were viewed at approximately the same time. Applying the subjective context principle *ItemHistory* is a feature that enables such an action. When working on a specific information item (e.g., a student working on a course paper), the user may open several other information items related to the same task (e.g., Web pages, e-mail, and other documents containing relevant information). In current design, the connection between these information items is lost, and the user needs to separately retrieve each of them. *ItemHistory* indexes these information items without bothering the user. When the need arises, the user will be able to use this feature (e.g., as an additional option in the item's File menu) to view and retrieve all items that were open at the same time as the current one (Figure 6). A similar history option also can be applied on a folder level, giving the user the ability to retrieve all information items that were open in parallel to the ones the folder contains (for previous work on use of personal information history, see Hill & Hollan, 1994; Hill, Hollan, Wroblewski, & McCandless, 1992).

Project contact list *(social context).* Projects often involve the exchange of information within a group of people that can be viewed as a subset of the user's contact list. In this feature, each folder has its own subset of contacts. The list is updated automatically with a person's contact information when an information item from that person (e.g., e-mail or attached document) is stored in that folder, or updated manually by dragging it from the user's general contact list (Figure 7). When e-mailing a document contained in the folder as an attachment, the user will see the names of people associated with the project at the top of the general list. This will not only allow for more efficient selection but also possibly remind the user of project-related people to whom the document should be sent.

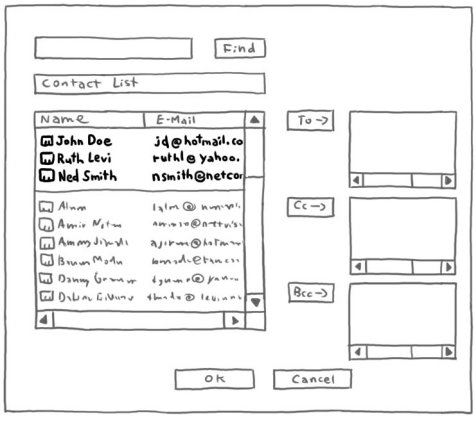


FIG. 7. Project Contact List scheme.


DOI: 10.1002/asi

Before concluding this article, we would like to address the way the user-subjective approach relates to contemporary directions in PIM design.

### *Hierarchical Design and Its Alternatives*

The current design of PIM systems is hierarchical: The system allows users to store their information items in folders (or directories), which in turn can contain subfolders (or subdirectories), and so forth. Lansdale (1988) criticized the use of hierarchies, and suggested three complementary alternative directions which have been extensively implemented in contemporary experimental systems:

- Multiple classification: Users are not restricted to storing an information item in a single folder. Each information item can be assigned to several categories or tags (Cutrell, Robbins, et al., 2006; Dourish et al., 1999; Kaplan, Kapor, Belove, Landsman, & Drake, 1990; Quan, Bakshi, Huynh, & Karger, 2003);
- Search: When searching for an information item, the users are not limited to specific folder categories. They can use any attribute they happen to remember about the information item to retrieve it. Strong advocates of the search tool believe that search will eliminate the need for hierarchical classification (Cutrell, Dumais, & Teevan, 2006; Dourish et al., 2000; Dumais et al., 2003; Fertig, Freeman, & Gelernter, 1996a, 1996b; Raskin, 2000);
- Automatic Classification: The system assigns or suggests a classification category for information items to take the burden off the user (Agrawal, Bayardo, & Srikant, 2000; Cutting, Karger, Pedersen, & Tukey, 1992; Whittaker & Sidner, 1996; Witten & Frank, 2005).

The user-subjective approach does not take a stand in the argument between hierarchical design and its alternatives, as all options for PIM design could benefit from systematic use of subjective attributes. Subjective attributes can be used as tags, assist in searches (as shown in the examples of *GrayArea* and *Old'nGray*), and serve as means for automatic classification. Thus, we encourage all PIM designers to use the user-subjective approach and its principles as generic ideas that call for applications and features to improve PIM system designs.

## Conclusions

The overall research findings indicate that personal computer users are interested in using subjective attributes, and indeed, use them whenever their system design encourages this; however, when system design discourages the use of subjective attributes, users avoid using them, or find alternative ways of doing so. Regarding classification, participants tended to save files of different formats together in project folders (when the design encouraged this), but tended not to store project-related e-mail and Web Favorites in these folders (as the design discouraged it). Regarding importance, participants tended to use the features that supported retrieval of information items of high importance; however, as the system does not offer a way of making items of low importance less visible, the users found their own alternative ways of doing so. As to the context, current design makes little use of this attribute; however, participants related to it when talking about their personal information organization. It seems that current systems do not offer comprehensive use of subjective attributes, and users could benefit from additional options that better exploit them.

These findings support the user-subjective approach and encourage the design of PIM systems according to its principles. We suggested seven novel implementations design schemes: *ProjectFolders*, *MyProjects*, and *ProjectField*, based on the subjective project classification principle; *GrayArea* and *Old'nGray*, in line with the subjective importance principle; and *ItemHistory* and *Project Contact List*, founded on the subjective context principle. Each of these design schemes should be developed and tested before being recommended for public usage; however, the wealth of straightforward novel implementations that the user-subjective approach suggests is an indication of its generative strength.

## Acknowledgments

We thank the participants for their time, Sharon Hardof-Yaffe and Karen Precel for helping us in collecting data, our anonymous reviewers for their eye-opening comments, and the Sacta Rashi Foundation for its part in funding the research.

## References

Abrams, D., Baecker, R., & Chignell, M. (1998). Information archiving with bookmarks: Personal Web space construction and organization. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 41–48). Los Angeles: ACM Press/Addison-Wesley.

Agrawal, R., Bayardo, R.J., & Srikant, R. (2000). Athena: Mining-based interactive management of text database. In Proceedings of the 7th International Conference on Extending Database Technology: Advances in Database Technology (pp. 365–379). Konstanz, Germany: Springer-Verlag.

Barreau, D.K. (1995). Context as a factor in personal information management systems. Journal of the American Society for Information Science, 46(5), 327–339.

Barreau, D.K., & Nardi, B.A. (1995). Finding and reminding: file organization from the desktop. SIGCHI Bulletin, 27(3), 39–43.

Bellotti, V., & Smith, I. (2000). Informing the design of an information management system with iterative fieldwork. In Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (pp. 227–237). New York: ACM Press.

Bergman, O., Beyth-Marom, R., & Nachmias, R. (2003). The user-subjective approach to personal information management systems. Journal of the American Society for Information Science and Technology, 54(9), 872–878.

Bergman, O., Beyth-Marom, R., & Nachmias, R. (2006). The project fragmentation problem in personal information management. In CHI 2006 Conference on Human Factors in Computing Systems (pp. 271–274). Montreal: ACM Press.

Bergman, O., Boardman, R., Gwizdka, J., & Jones, W.P. (2004). Personal information management. In CHI '04 Extended Abstracts on Human Factors in Computing Systems (pp. 1598–1599). Vienna: ACM Press.

Boardman, R., & Sasse, M.A. (2004). "Stuff goes into the computer and doesn't come out:" A cross-tool study of personal information management. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 583–590). Vienna: ACM Press.


DOI: 10.1002/asi

Boardman, R., Spence, R., & Sasse, M.A. (2003). Too many hierarchies? The daily struggle for control of the workspace. In J. Jacko & C. Stephanidis (Eds.), HCI International Crete, Greece, 2003 (Vol. 1, pp. 616–620). Mahwah, NJ: Erlbaum.

Bush, V. (1945). As we may think. The Atlantic Monthly, 176(1), 101–108.

Chalmers, M. (2004). A historical view of context. Computer Supported Cooperative Work, 13(3), 223–247.

Cutrell, E., Dumais, S.T., & Teevan, J. (2006). Searching to eliminate personal information management. Communications of the ACM, 49(1), 58–64.

Cutrell, E., Robbins, D.C., Dumais, S.T., & Sarin, R. (2006). Fast, flexible filtering with Phlat: Personal search and organization made easy. In CHI 2006 Conference on Human Factors in Computing Systems (pp. 261–270). Montreal: ACM Press.

Cutting, D.R., Karger, D.R., Pedersen, J.O., & Tukey, J.W. (1992). Scatter/gather: A cluster-based approach to browsing large document collections In Proceedings of the 15th annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 318–329). Copenhagen: ACM Press.

Dey, A.K. (2001). Understanding and using context. Personal and Ubiquitous Computing, 5,(1), 4–7.

Dourish, P. (2004). What we talk about when we talk about context. Personal and Ubiquitous Computing, 8(1), 19–30.

Dourish, P., Edwards, W.K., LaMarca, A., Lamping, J., Petersen, K., Salisbury, M., et al. (2000). Extending document management systems with user-specific active properties. ACM Transactions on Information Systems, 18(2), 140–170.

Dourish, P., Edwards, W.K., LaMarca, A., & Salisbury, M. (1999). Presto: An experimental architecture for fluid interactive document spaces. ACM Transactions on Computer–Human Interactions, 6(2), 133–161.

Dumais, S.T., Cutrell, E., Cadiz, J.J., Jancke, G., Sarin, R., & Robbins, D.C. (2003). Stuff I've seen: A system for personal information retrieval and re-use. In Proceedings of the 26th annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 72–79). Toronto: ACM Press.

Fertig, S., Freeman, E., & Gelernter, D. (1996a). "Finding and reminding" reconsidered. SIGCHI Bulletin, 28(1), 66–69.

Fertig, S., Freeman, E., & Gelernter, D. (1996b). Lifestreams: An alternative to the desktop metaphor. In Conference Companion on Human Factors in Computing Systems: Common Ground (pp. 410–411). Vancouver, British Columbia, Canada: ACM Press.

Freeman, E., & Gelernter, D. (1996). Lifestreams: A storage model for personal data. SIGMOD Record, 25(1), 80–86.

Hill, W.C., & Hollan, J.D. (1994). History-enriched digital objects: Prototypes and policy issues. The Information Society, 10(2), 139–145.

Hill, W.C., Hollan, J.D., Wroblewski, D., & McCandless, T. (1992). Edit wear and read wear. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Monterey, CA: ACM Press.

Jones, W., & Bruce, H. (2005). A report on the NSF-Sponsored Workshop on Personal Information Management. Seattle, WA: National Science Foundation.

Jones, W., Bruce, H., & Dumais, S. (2001). Keeping found things found on the web. In Proceedings of the 10th International Conference on Information and Knowledge Management (pp. 119–126). Atlanta, GA: ACM Press.

Jones, W., Karger, D., Bergman, O., Franklin, M., Pratt, W., & Bates, M. (2005). Towards a unification & integration of PIM support. Seattle, WA: National Science Foundation.

Jones, W., Munat, C.F., & Bruce, H. (2005). The universal labeler: Plan the project and let your information follow. In A. Grove (Ed.), 68th Annual Meeting of the American Society for Information Science and Technology (ASIST 2005) (Vol. 42, pp. TBD.). Charlotte, NC: American Society for Information Science & Technology.

Jones, W., Phuwanartnurak, A.J., Gill, R., & Bruce, H. (2005). Don't take my folders away! Organizing personal information to get things done. In CHI '05 Extended Abstracts on Human Factors in Computing Systems. Portland, OR: ACM Press.

Jones, W.P. (2004). Finders, keepers? The present and future perfect in support of personal information management. First Monday, 9(3).

Kaplan, S.J., Kapor, M.D., Belove, E.J., Landsman, R.A., & Drake, T.R. (1990). Agenda: A personal information manager. Communications of the ACM, 33(7), 105–116.

Kaptelinin, V. (2003). UMEA: Translating interaction histories into project contexts. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 353–360). Ft. Lauderdale, FL: ACM Press.

Karger, D.R., & Quan, D. (2004). Collections: Flexible, essential tools for information management. In CHI '04 Extended Abstracts on Human Factors in Computing Systems. Vienna: ACM Press.

Kwasnik, B.H. (1991). The importance of factors that are not document attributes in the organization of personal documents. Journal of Documentation, 47, 389–398.

Lansdale, M.W. (1988). The psychology of personal information management. Applied Ergonomics, 19(1), 55–66.

Malone, T.W. (1983). How do people organize their desks? Implications for the design of office information systems. ACM Transactions on Office Information Systems, 1, 99–112.

Medlin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. Psychological Review, 85, 207–238.

Quan, D., Bakshi, K., Huynh, D., & Karger, D.R. (2003). User interfaces for supporting multiple categorization. In M. Rauterberg (Ed.), Proceedings of INTERACT 2003 (pp. 228–235). Amsterdam: IOS Press.

Raskin, J. (2000). The humane interface: New directions for designing interactive systems. Boston: ACM Press/Addison-Wesley.

Saracevic, T. (1999). Information science. Journal of the American Society for Information Science, 50(12), 1051–1063.

Schamber, L. (1994). Relevance and information behavior. Annual Review of Information Science and Technology, 29, 3–48.

Soergel, D. (1985). Organizing information: Principles of data base and retrieval systems. San Diego, CA: Academic Press.

Taylor, R.S. (1986). Value added processes in information systems. Norwood, NJ: Ablex.

Teevan, J., Jones, W.P., & Bederson, B.B. (2006). Personal information management. Communications of the ACM, 49(1), 40–43.

Whittaker, S., & Sidner, C. (1996). Email overload: Exploring personal information management of email. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Common Ground (pp. 276–283). Vancouver, British Columbia, Canada: ACM Press.

Whittaker, S., Terveen, L., & Nardi, B.A. (2000). Let's stop pushing the envelope and start addressing it: A reference task agenda for HCI. Human–Computer Interaction, 15, 75–106.

Witten, I.H., & Frank, E. (2005). Data mining: Practical machine learning tools and techniques (2nd ed.). San Francisco: Kaufmann.


DOI: 10.1002/asi